

# Using Administrative Records

Andrew A. Samwick

Census Scientific Advisory Committee Fall 2018 Meeting

December 7, 2018

# What Might We Strive For?

- Perhaps, a data universe in which:
  - The only questions asked to respondents directly are the ones that cannot be answered accurately by data already available through non-survey methods.
  - The burden on respondents is minimized in answering the questions that they are asked directly.
  - Census datasets have the coverage, frequency, and depth needed to inform research, policy-making, and worthwhile private pursuits.
- Administrative records can help with each of these three objectives.

# Possible Concerns

- The gold standard of quality that is the hallmark of Census data:
  - Coverage and representativeness
  - Alignment and accuracy
  - Confidentiality protection

# Possible Concerns

- Incorporating AR into the data product **should** raise costs in the short term and lower them in the long term.
- Validating the use of AR as a replacement requires running the existing methodology and the alternative methodology in parallel until we can demonstrate acceptable quality going forward.
- Perhaps some of this validation will occur gradually as AR are used for editing missing data in existing surveys.

# One Additional Dimension

- Are we focused on the PUMS of individual data or the aggregates by geography?
- The scope for using AR from third-party providers is likely greater the larger is the geographic unit of analysis.

# Using AR to the Greatest Effect

- As in the ACS example, look for characteristics that:
  - May not be apparent to the respondent during a survey,
  - Don't change frequently, and
  - When they change, the amount of the change is already tracked by some other data collection entity.
- So the triage on ACS Housing Unit data is a good first step.
  - Focus on Acreage, Year Built, Property Value, and Real Estate Taxes
  - CSAC would be happy to see the working paper on the validation exercise.

# The NES-D Project

- At our Spring [meeting](#), we learned that the [Annual Business Survey](#) would replace the [Survey of Business Owners](#), and some other surveys, and move to an annual frequency.
- However, the ABS excludes non-employer businesses, leaving us without a comprehensive source of owner demographic data for non-employer business.
- The [Nonemployer Statistics](#) is a separate, annual survey for businesses that have no paid employees and are subject to federal income tax. This series includes only the number of businesses and total receipts by industry.

# NES-D Challenges

- Use the NES as the sampling frame for non-employer firms, and use primarily the ACS to impute race, gender, age, place of birth, and Hispanic origin.
- The assignment of AR demographics will only be as good as the external sources of data used for imputation and the ability to link observations via Protected Identification Keys (PIKs).
- As with the ACS efforts, CSAC would be happy to see the working paper on the validation exercises.



# Look Very Broadly for Collaborators

- The Census Bureau is not the only entity that would benefit from better administrative data.
  - For example, consider how many redundant purchases of title insurance there are in real estate transactions when a property is resold.
  - Eliminating redundancies through an authentic, public-use dataset could generate a lot of surplus to be shared.

## (Remaining) Questions from Adela to CSAC

- AR data allows for accuracy improvements (e.g. number of owners to assign firm-level demographics) which can create breaks with historical time series: what is the role of stakeholders and census in balancing this tension? How should we balance accuracy vs consistency?
- What is best way to build support for NES-D and ensuring external awareness?

## (Remaining) Questions from Jennifer to CSAC

- The list of research initiatives is long and continues to grow. Does the committee have opinions on what they would like to see us focus on first?
- Are there other research initiatives regarding administrative data that we should be aware of?

Thank You!